A SYNOPSIS ON

"Design and Development of Models for Early Detection of Breast Cancer using Soft Computing Techniques"

Submitted to

Swami Ramanand Teerth Marathwada University, Nanded

for the Degree of

Doctor of Philosophy in Computer Science

in

Faculty of Science and Technology

by

Mrs. Kavita P. Shinde

Assistant Professor

MIT Arts, Commerce and Science College, Alandi (D), Pune.

Under the Guidance of

Dr. U. V. Kulkarni

Professor

Department of Computer Science and Engineering

Shri Guru Gobind Singhji Institute of Engineering & Technology, Nanded

2021-2022

INDEX

Sr. No.	Title	Page No.
1.	Abstract	1
2.	Introduction	1-2
3.	The Review of Literature	2-7
4.	Objectives of the Study	7
5.	Research Methodology	7
6.	Significance of Study	8
7.	Proposed Work Plan	8
8.	References	8-10

1. Abstract:

Breast cancer is most popular and growing disease in the world. Breast cancer is mostly found in the women. Early detection is a way to control the breast cancer [17]. Breast cancer is a devastating disease, with high mortality rates around the world. Early detection of breast cancer through screening tests such as mammograms is an efficient way to maximize patients' survival rate by treating the disease prematurely [16].

Creating a model which will automatically detect breast cancer in early stage is challenging task. Many machine learning algorithms such as K-Nearest neighbors (KNN), Decision Tree (DT), Naive Bayes (NB) and Support Vector Machine (SVM) have showed significant improvement compared to artificial intelligence but these algorithms require data to fed hand crafted. A deep learning algorithm learns these features on its own from data. So, deep learning algorithms such as ResNet, GoogleNet and MobileNet showed promising improvement in results with the help of transfer learning and regularization techniques.

The motivation behind this research is to design and develop models for early detection of breast cancer using soft computing techniques which will prevent late treatments due to false negatives as well as preventing unnecessary treatments in cases of false positives.

2. Introduction:

Breast and lung cancers were the most common cancers worldwide, contributing 12.5% and 12.2% of the total number of new cases diagnosed in 2020. Colo-rectal cancer was the third most common cancer with 1.9 million new cases in 2020, contributing 10.7% of new cases [1]. The International Agency for Research on Cancer (IARC) estimates that globally, 1 in 5 people develop cancer during their lifetime, and 1 in 8 men and 1 in 11 women die from the disease. According to survey, as shown in fig, (1), for the 1st time, female breast cancer has become the most commonly diagnosed cancer and it is also the primary cause of mortality due to cancer in female around the world [2]. Cancer occurs when changes called mutations take place in genes that regulate cell growth. The mutations let the cells divide and multiply in an uncontrolled, chaotic way. Breast cancer occurs when a malignant (cancerous) tumor originates in the breast. As breast cancer tumors mature, they may metastasize (spread) to other parts of the body [8].

Breast cancer tumors can be categorized into two broad scenarios.

- Benign (Noncancerous): Benign cases are considered as noncancerous, that is, non-life-threatening. But on a few occasions, it could turn into a cancer status [5].

- Malignant (Cancerous): Malignant cancer starts from an abnormal cell growth and might rapidly spread or invade nearby tissue. Normally the nuclei of the malignant tissue are much bigger than in normal tissue, which can be life-threatening in future stages [5].

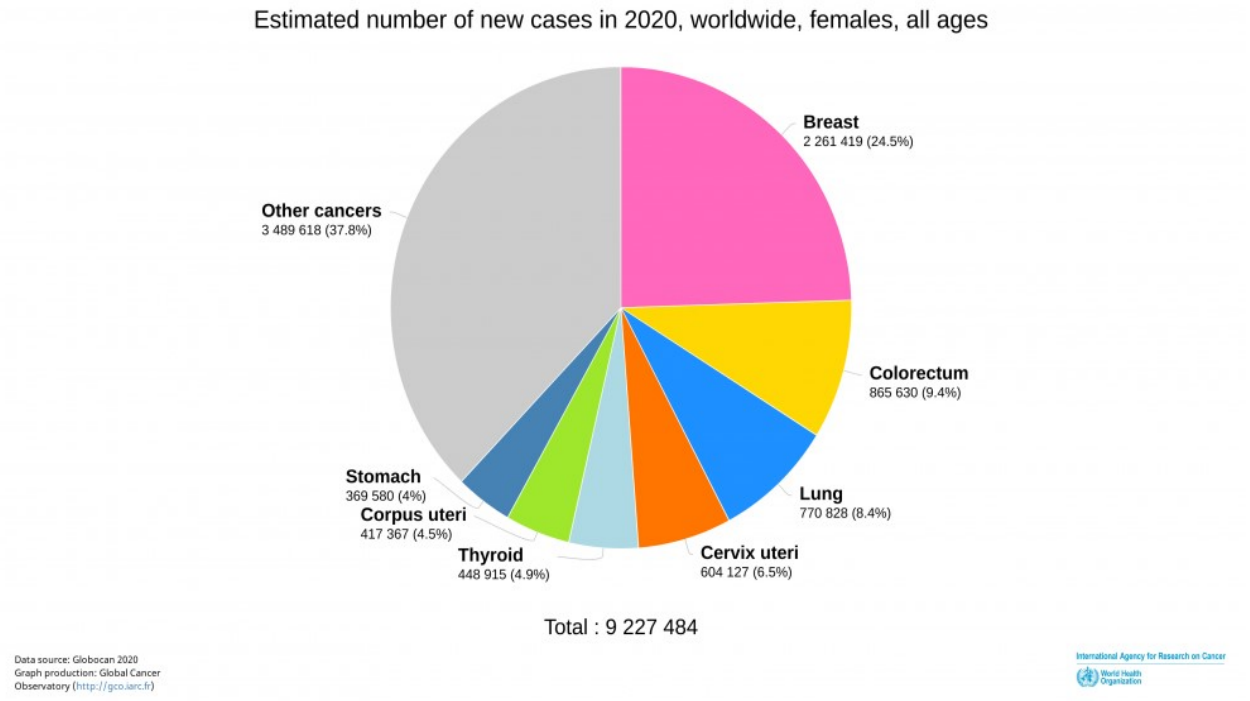

Fig. 1: GLOBOCAN 2020: New Global Cancer Data (2)

3. The Review of Literature:

In this section, some of the related works previously done on breast cancer diagnosis by researchers using different machine learning and deep learning algorithms are discussed and their results are compared.

-Vaka, A. R., Soni, B., & K., S. R. (2020) [1]. In this paper, the author has studied performance analysis of most popular breast cancer detection algorithms like NB classifier, SVM classifier, Bi-clustering and Ada boost techniques, RCNN classifier and Bidirectional Recurrent Neural Networks (HA-BiRNN) and found that all algorithms have some disadvantages. The authors presented the new method DNNS for detecting breast cancer. The proposed method is based on support value on a deep neural network. To meet the better performance, efficiency, and quality of

images, a normalization process has been employed. Experimental results proved that the proposed DNNS is quite better than the existing methods. It is ensured that the proposed algorithm is advantageous in both performance, efficiency and qualities of images are crucial in the latest medical systems.

-Aswathy, M. A., & Jagannath, M. (2020) [2]. In this paper, the author carried out performance analysis of segmentation algorithms for the detection of breast cancer. The study used a publicly available data set UCSB, with normal (benign) and abnormal (malignant) images. It consists of three different stages image enhancement, nuclei segmentation, extraction of features and classification. The segmentation has been done using three algorithms- fuzzy c-means clustering, k-means clustering and ACM. Among these, k-means clustering has the maximum segmentation accuracy of 93%. Other performance metrics are also superior for k-means clustering algorithm. The automated classification of breast cancer images is based on SVM model on GLCM features and achieved an accuracy of 91.1%. The SVM classification approach suggested in this study permits to achieve high sensitivity also.

-Ara, S., Das, A., & Dey, A. (2021) [3]. In this paper, Wisconsin Breast Cancer Data set (WBCD) of the FNA biopsy system has been used and different machine learning classifiers have been implemented to determine the form of breast cancer in a suspected patient. Six classification models were used including random forest, logistic regression, decision tree, naive bayes, SVM and KNN. In order to discover the most suited model for predicting breast cancer, the results acquired are then evaluated to compare the algorithms. Maximum accuracy is achieved using Random Forest and SVM on the test set which is 96.5%. Their training accuracy is also higher than other algorithms.

-Alanazi, S. A., Kamruzzaman, M. M., Islam Sarker, M. N., Alruwaili, M., Alhwaiti, Y., Alshammari, N., & Siddiqi, M. H. (2021) [4]. In this paper, a CNN method is proposed in this study to boost the automatic identification of breast cancer by analyzing hostile ductal carcinoma tissue zones in whole-slide images (WSIs). This paper investigates the proposed system that uses various CNN architectures to automatically detect breast cancer, comparing the results with those from machine learning algorithms. All architectures were guided by a big data set of about 275,000, 50 × 50-pixel RGB image patches. Validation tests were done for quantitative results using the performance measures for every methodology. The proposed system is found to be successful, achieving results with 87% accuracy, which could reduce human mistakes in the diagnosis process.

Moreover, the proposed system achieves accuracy higher than the 78% accuracy of machine learning algorithms. Therefore, this improves accuracy by 9% above results from machine learning algorithms.

-Zhang, Z., Chen, B., Xu, S., Chen, G., & Xie, J. (2021) [5]. In this paper, a novel V-CDNN model has been proposed for diagnosing breast cancer. Voting strategy has been applied to make diagnosis results more accurate and convincing in the testing process. Experimental results showed that V-CDNN has greatly improved computation efficiency and increased accuracy rate of diagnosis by adopting neural dynamic learning algorithm and training less parameters than traditional gradient-based neural networks. The accuracy rate has reached 100% average diagnosis accuracy.

-Saber, A., Sakr, M., Abo-Seida, O. M., Keshk, A., & Chen, H. (2021) [6]. In this paper, author developed a new deep learning model based on the transfer-learning technique for automatic detection and diagnosis of the breast cancer suspected area. Deep learning architectures are modeled to be problem-specific. Features are extracted from the mammography image analysis-society (MIAS) data set using a pre-trained CNN architecture such as Inception V3, ResNet50, Visual Geometry Group networks (VGG)-19, VGG-16, and Inception-V2 ResNet. Six evaluation metrics such as accuracy, sensitivity, specificity, precision, f1-score, and area under the ROC curve (AUC) has been chosen for performance evaluation. Experimental results show that the transfer learning of the VGG16 model is powerful for breast cancer diagnosis by classifying the mammogram breast images with overall accuracy, sensitivity, specificity, precision, f1-score, and AUC of 98.96%, 97.83%, 99.13%, 97.35%, 97.66%, and 0.995, respectively for 80-20 method and 98.87%, 97.27%, 98.2%, 98.84%, 98.04%, and 0.993 for 10-fold cross-validation method.

-Chugh, G., Kumar, S., & Singh, N. (2021) [7]. In this paper, author reviewed peer-reviewed, newly developed computer aided diagnosis (CAD) systems to implement machine learning and deep learning techniques for diagnosing breast carcinoma. Authors also compare them with previously established methods and provide technical details with the pros and cons for each model. They also discussed open issues, research gaps, and future research directions for the advanced CAD models in medical image analysis.

-Koh, J., Yoon, Y., Kim, S., Han, K., & Kim, E. K. (2022) [8]. In this paper, the author has proved that using deep learning, the sensitivity of incidental breast cancer detection on chest CT would

improve. Results were validated in the internal and external datasets. Study collected 1170 preoperative chest CT scans after the diagnosis of breast cancer. For algorithm development (n = 1070), internal test (n = 100), and external test (n = 100). A deep learning algorithm based on RetinaNet was developed and tested to detect breast cancer on chest CT. Results: In the internal test set, the algorithm detected 96.5% of breast cancers with 13.5 false positives per case (FPs/case). In the external test set, the algorithm detected 96.1% of breast cancers with 15.6 FPs/case. When the candidate probability of 0.3 was used as the cutoff value, the sensitivities were 92.0% with 7.36 FPs/case for the internal test set and 93.0% with 8.85 FPs/case for the external test set. When the candidate probability of 0.4 was used as the cutoff value, the sensitivities were 88.5% with 5.24 FPs/case in the internal test set and 90.7% with 6.3 FPs/case in the external test set.

-Dewangan, K. K., Dewangan, D. K., Sahu, S. P., & Janghel, R. (2022) [9]. In this paper, a novel Back Propagation Boosting Recurrent Wienmed model (BPBRW) with Hybrid Krill Herd African Buffalo Optimization (HKH-ABO) mechanism is developed for detecting breast cancer in an earlier stage using breast MRI images. MRI breast images are initially trained to the system, and an innovative Wienmed filter is established for prepossessing the MRI noisy image content. BPBRW with HKH-ABO mechanism also categorizes the breast cancer tumor as benign and malignant. This novel algorithm has achieved 99.6% higher accuracy with a 0.12% lower error rate.

-Yu, X., Zhou, Q., Wang, S., & Zhang, Y. D. (2022) [10]. In this paper, authors reviewed previous and current work on CAD systems for breast cancer. Focus was on detection, segmentation, and classification of mass and micro calcification. Comparison between conventional CAD and deep learning-based CAD is done in this paper. CNN's are widely used in solving classification and detection problems; FCNN's are generally deployed for segmentation tasks. However, we found that segmentation works regarding mass significantly outnumbered the works on calcification as mass detection and classification tend to more interesting in the area. Authors concluded that, deep-learning-based methods have become a predominating choice in the implementation of breast cancer CAD systems.

-Zhang, Z., Li, Y., Wu, W., Chen, H., Cheng, L., & Wang, S. (2021) [11]. In this paper, accurate and efficient breast tumor detection of automated breast ultrasound (ABUS) plays a significant role for clinical diagnosis. However, reviewing thousands of ABUS slices is extremely a time-consuming task. To improve the tumor detection rate and reduce FP regions, a Bayesian YOLOv4 network based on Monte Carlo dropout (MC Drop) is proposed in this paper. Firstly, some tricks

in YOLOv4 network are adjusted to make it more suitable for ABUS tumor detection. Secondly, MC-Drop is applied to introduce uncertainty into the detection network and the Bayesian YOLOv4 network is constructed, which effectively reduces accidental FP regions and improves the detection rate for hard tumor region. This method was tested on 68 volumes, including 21,624 slices (1683 tumor slices and 19,941 normal slices). It obtains a promising result with sensitivity of 0.88 and FPs/S at 0.19.

-Assegie, T. A. (2021) [12]. In this research, an optimized KNN model is proposed for breast cancer prediction using a grid search approach for searching the best hyper-parameter. A comparative study is carried out between default hyper-parameter and tuned hyper-parameter performance. Result shows that the performance significantly improves when best hyper parameter or K value is used for training the KNN. The performance of the KNN with default parameters is 90.10%. Better breast cancer detection accuracy is achieved by the KNN using best hyper-parameter or K value is chosen using a grid search approach when the algorithm is trained and the highest performance achieved using hyper-parameter tuning is 94.35%.

-Mahmood, T., Li, J., Pei, Y., Akhtar, F., Imran, A., & Ur Rehman, K. (2020) [13]. In this paper, authors systematically compared the strengths, limitations, and performance of recent deep learning and machine learning schemes by analyzing medical multi-image modalities. CAD system to limit the computational and time complexities related to the development of breast cancer diagnostic system. Machine learning approaches are found imperfect for precise segmentation of densities; however, deep learning approaches helped in minimizing false- positive ratio (FPR) in the segmentation of masses. The review showed that with the advancement of deep learning approaches the process of breast abnormalities segmentation and classification is improved.

-Sharma, V., Singh, M., Galgotias [14]. In this paper, authors used CNN for diagnosis breast cancer and also implemented same data set on other machine learning algorithm such as neural network, SVM, and RF. Authors used Wisconsin breast cancer data set which is publicly available for researcher. This database is generated from biopsy images that having 569 rows and 30 columns. First, they have pre-processed the data set and extracted useful feature using RFECV model then implemented encoder and found final training data set. For the trained the data set they have used deep learning model convolutional neural network algorithm and achieved 99.67% accuracy.

-Eskreis-Winkler, S., Onishi, N., Pinker, K., Reiner, J. S., Kaplan, J., Morris, E. A., & Sutton, E. J. (2021) [15]. In this paper, authors showed that thermography imaging is an effective diagnostic technique which is used for breast cancer detection with the help of infrared technology. They proposed a fully automatic breast cancer detection system. First, U-Net network is used to automatically extract and isolate the breast area from the rest of the body which behaves as noise during the breast cancer detection model. Second, they propose a two-class deep learning model, which is trained from scratch for the classification of normal and abnormal breast tissues from thermal images. The proposed system is evaluated using real data (A benchmark, database (DMR-IR)) and achieved accuracy = 99.33%, sensitivity = 100% and specificity = 98.67%.

-Jaamour Supervised by David Harris-Birtill, A., & McMillan, L. (n.d.) [16]. In this paper, author designed and implemented a deep learning pipeline capable of detecting cases of breast cancer in mammograms through various deep learning techniques. After investigating a wide array of techniques using a bag-of-tricks approach, a fully-functional pipeline with data preprocessing steps, a CNN model for learning the data and prediction visualization was created. The most positive result (67.08%) on CBIS-DDSM came from transfer learning techniques, using Image Net weights with MobileNetV2 and binary mini-MIAS weights for the custom MLP layers, coupled with class weight techniques for balancing the datasets.

4. Objectives of the Study:

- Extracting and isolating the breast area automatically from other parts of mammography images.
- Proposing a model for automatically classifying benign and malignant from given data set.
- Evaluating the performance of the proposed model using accuracy, sensitivity and specificity.
- Explore techniques that can accurately detect breast cancer in order to prevent late treatments due to false negatives as well as preventing unnecessary treatments in cases of false positives.

5. Research Methodology:

The deep learning strategy for implementation of breast cancer detection can be broken down in following phases:

- **Datasets:** The first step in this research is to choose data set. Different datasets are available for prediction of breast cancer. Ex: MIAS, RIDER Breast MRI, in breast Etc.
- **Data Preprocessing**: Next task is to load a datasets in memory and label images for the classification task.
- **Model training:** Creating a CNN model that can fit the data to learn the training set samples. Predictions are carried out once the model finishes training on the validation and test sets.
- **Result visualization:** The model's performance is evaluated by calculating various metrics and plotting predictions.

6. Significance of Study:

The main aim of my research is to automate methods to identify breast cancer in early stage using deep learning algorithms with the help of transfer learning and regularization techniques to improve the operational efficiency of radiologists and trained professionals so that mortality rate of women decreases.

7. Proposed Work Plan:

Year	Plan
1.	Literature Review & Basic Implementation
2.	Design and Development of Model
	Data set Selection
	Data P reprocessing
	Model Training
	Result Visualization
3.	Research Paper Writing and Publishing
4.	Thesis Writing

8. References :

1. Vaka, A. R., Soni, B., & K., S. R. (2020). Breast cancer detection by leveraging Machine Learning. ICT Express, 6(4), 320–324. https://doi.org/10.1016/j.icte.2020.04.009

2. Aswathy, M. A., & Jagannath, M. (2020). Performance Analysis of Segmentation Algorithms for the Detection of Breast Cancer. Procedia Computer Science, 167, 666–676. https://doi.org/10.1016/j.procs.2020.03.333.
3. Ara, S., Das, A., & Dey, A. (2021). Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms. 2021 International Conference on Artificial Intelligence, ICAI 2021, 97–101. https://doi.org/10.1109/ICAI52203.2021.9445249.
4. Alanazi, S. A., Kamruzzaman, M. M., Islam Sarker, M. N., Alruwaili, M., Alhwaiti, Y., Alshammari, N., & Siddiqi, M. H. (2021). Boosting Breast Cancer Detection Using Convolutional Neural Network. Journal of Healthcare Engineering, 2021. https://doi.org/10.1155/2021/5528622
5. Zhang, Z., Chen, B., Xu, S., Chen, G., & Xie, J. (2021). A novel voting convergent difference neural network for diagnosing breast cancer. Neurocomputing, 437, 339–350. https://doi.org/10.1016/j.neucom.2021.01.083
6. Saber, A., Sakr, M., Abo-Seida, O. M., Keshk, A., & Chen, H. (2021). A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique. IEEE Access, 9, 71194–71209. https://doi.org/10.1109/ACCESS.2021.3079204
7. Chugh, G., Kumar, S., & Singh, N. (2021). Survey on Machine Learning and Deep Learning Applications in Breast Cancer Diagnosis. Cognitive Computation, 13(6), 1451–1470. https://doi.org/10.1007/s12559-020-09813-6. Springer Science+Business Media, LLC part of Springer Nature 2021
8. Koh, J., Yoon, Y., Kim, S., Han, K., & Kim, E. K. (2022). Deep Learning for the Detection of Breast Cancers on Chest Computed Tomography. Clinical Breast Cancer, 22(1), 26–31. https://doi.org/10.1016/j.clbc.2021.04.015
9. Dewangan, K. K., Dewangan, D. K., Sahu, S. P., & Janghel, R. (2022). Breast cancer diagnosis in an early stage using novel deep learning with hybrid optimization technique. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-022-12385-2.
10. Yu, X., Zhou, Q., Wang, S., & Zhang, Y. D. (2022). A systematic survey of deep learning in breast cancer. International Journal of Intelligent Systems, 37(1), 152–216. https://doi.org/10.1002/int.22622
11. Zhang, Z., Li, Y., Wu, W., Chen, H., Cheng, L., & Wang, S. (2021). Tumor detection using deep learning method in automated breast ultrasound. Biomedical Signal Processing and Control, 68. https://doi.org/10.1016/j.bspc.2021.102677

12. Assegie, T. A. (2021). An optimized K-Nearest neighbor based breast cancer detection. Journal of Robotics and Control (JRC), 2(3), 115–118. https://doi.org/10.18196/jrc.2363
13. Mahmood, T., Li, J., Pei, Y., Akhtar, F., Imran, A., & Ur Rehman, K. (2020). A brief survey on breast cancer diagnostic with deep learning schemes using multi-image modalities. IEEE Access, 8, 165779–165809. https://doi.org/10.1109/ACCESS.2020.3021343
14. Sharma, V., Singh, M., Galgotias .Breast Cancer Diagnosis Using Deep Learning Algorithm. Proceedings, IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking : (ICACCCN) : on 12th-13th Oct, 2018.
15. Eskreis-Winkler, S., Onishi, N., Pinker, K., Reiner, J. S., Kaplan, J., Morris, E. A., & Sutton, E. J. (2021). Using Deep Learning to Improve Nonsystematic Viewing of Breast Cancer on MRI. Journal of Breast Imaging, 3(2), 201–207. https://doi.org/10.1093/jbi/wbaa102
16. Jaamour Supervised by David Harris-Birtill, A., & McMillan, L. (n.d.). Breast Cancer Detection in Mammograms using Deep Learning Techniques.
17. Priyanka, & Sanjeev, K. (2021). A review paper on breast cancer detection using deep learning. IOP Conference Series: Materials Science and Engineering, 1022(1). https://doi.org/10.1088/1757-899X/1022/1/012071.

Websites:

1. https://www.wcrf.org/dietandcancer/worldwide-cancer-data/
2. https://www.uicc.org/news/globocan-2020-new-global-cancer-data
3. https://towardsdatascience.com/building-a-simple-machine-learning-model-on-breast-cancer-data-eca4b3b99fa3.
4. https://healthitanalytics.com/news/deep-learning-may-detect-breast-cancer-earlier-than-radiologists#:~:text=January%2014%2C%202021%20%2D%20A%20deep,study%20published%20in%20Nature%20Medicine.

Mrs. Kavita P. Shinde
Research Student

Dr. U.V. Kulkarni
Research Guide

Place: Nanded

Date: